



# Can Automated Feedback Improve Teachers' Uptake of Student Ideas? Evidence From a Randomized Controlled Trial In a Large-Scale Online Course

**Dorottya Demszky**  
Stanford University

**Jing Liu**  
University of Maryland

**Heather C. Hill**  
Harvard University

**Dan Jurafsky**  
Stanford University

**Chris Piech**  
Stanford University

Providing consistent, individualized feedback to teachers is essential for improving instruction but can be prohibitively resource intensive in most educational contexts. We develop an automated tool based on natural language processing to give teachers feedback on their uptake of student contributions, a high-leverage teaching practice that supports dialogic instruction and makes students feel heard. We conduct a randomized controlled trial as part of an online computer science course, Code in Place (n=1,136 instructors), to evaluate the effectiveness of the feedback tool. We find that the tool improves instructors' uptake of student contributions by 24% and present suggestive evidence that our tool also improves students' satisfaction with the course. These results demonstrate the promise of our tool to complement existing efforts in teachers' professional development.

VERSION: November 2021

Suggested citation: Demszky, Dorottya, Jing Liu, Heather C. Hill, Dan Jurafsky, and Chris Piech. (2021). Can Automated Feedback Improve Teachers' Uptake of Student Ideas? Evidence From a Randomized Controlled Trial In a Large-Scale Online Course. (EdWorkingPaper: 21-483). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/thn9-wh86>

# Can Automated Feedback Improve Teachers' Uptake of Student Ideas?

## Evidence From a Randomized Controlled Trial In a Large-Scale Online Course

Dorottya Demszky<sup>1</sup>, Jing Liu<sup>2</sup>, Heather C. Hill<sup>3</sup>, Dan Jurafsky<sup>1</sup>, and Chris Piech<sup>1</sup>

<sup>1</sup> Stanford University

<sup>2</sup> University of Maryland

<sup>3</sup> Harvard University

### Abstract

Providing consistent, individualized feedback to teachers is essential for improving instruction but can be prohibitively resource-intensive in most educational contexts. We develop an automated tool based on natural language processing to give teachers feedback on their uptake of student contributions, a high-leverage teaching practice that supports dialogic instruction and makes students feel heard. We conduct a randomized controlled trial as part of an online computer science course, Code in Place (n=1,136 instructors), to evaluate the effectiveness of the feedback tool. We find that the tool improves instructors' uptake of student contributions by 24% and present suggestive evidence that our tool also improves students' satisfaction with the course. These results demonstrate the promise of our tool to complement existing efforts in teachers' professional development.

---

**Acknowledgements:** We are grateful to Michael Chang, Brahm Capoor, Julie Zelenski and other members of the Code in Place team for their help with implementing the project. We are also thankful to Greg Walton, Betty Malen, Gábor Orosz, János Perczel, Christine Kuzdzal, Grace Hu, Sterling Alic and the attendants of seminars at Brown, Google, Cornell and Stanford for their feedback. The project was partially funded by the HAI Hoffman-Yee grant (#203116). Demszky acknowledges support of the Melvin and Joan Lane Stanford Graduate Fellowship.

Dorottya Demszky ([ddemszky@stanford.edu](mailto:ddemszky@stanford.edu)) is a doctoral candidate in Linguistics at Stanford University. Jing Liu is assistant professor in Education Policy at the University of Maryland and a research affiliate at the IZA Institute of Labor Economics. Heather Hill is the Jerome T. Murphy professor in Education at the Harvard Graduate School of Education. Dan Jurafsky is professor of Linguistics and professor of Computer Science at Stanford University. Chris Piech is assistant professor of Computer Science Education at Stanford University.

## Introduction

A growing body of literature suggests that formative feedback is key to improving teachers' instruction and students' academic achievement (Taylor & Tyler, 2012; Steinberg & Sartain, 2015). In recent years, the contexts in which teaching and learning take place have become more diverse; these contexts now include online learning (e.g., virtual academies, MOOCs) and other informal settings (e.g., tutoring, after-school programs). However, in many of these settings, opportunities for teachers to receive feedback to improve their instruction are limited. Even preK-12 schooling, perhaps the most resource-rich environment for instructor learning, typically offers teachers only a few hours a year of professional development (PD), and even fewer opportunities for direct observation and feedback on practice (Cohen & Goldhaber, 2016).

There are two key challenges to developing scalable feedback mechanisms to improve teachers' instruction. First, generating consistent and individualized feedback tends to be *resource-intensive* because it requires classroom observation by instructional experts. For example, in U.S. K-12 classrooms, a pedagogical expert such as a principal, coach or mentor teacher might use the district protocol or an observation instrument such as the Classroom Assessment Scoring System (CLASS) (My Teaching Partner; Gregory et al., 2017) to observe and provide feedback to instructors several times a year (Cohen & Goldhaber, 2016). While such traditional forms of PD can be valuable, the required pedagogical expertise is often unavailable in non-preK-12 teaching contexts and, in K-12 settings, limited by financial constraints and teachers' unwillingness to allow outsiders to view their practice (Russell et al., 2020).

The second key challenge lies in developing solutions that are *effective* in improving teacher practices. In a recent STEM meta-analysis, less than half of math and science PD program impact estimates showed positive effects on teacher knowledge and practice; only one third showed positive impacts on student outcomes (Lynch et al., 2019). A related review of 76 IES-funded studies found that 36% of the interventions had no positive impact on teacher practice (Hill & Erickson, 2019). Even resource-intensive and successful PD programs often make only marginal changes in teachers' practice (Ball & Cohen, 1999; Borko, 2004; Garet et al., 2008, 2010, 2011; Jacob & McGovern, 2015). For instance, after up to 118 hours in a middle

school mathematics PD program that targeted teachers’ content knowledge and standards-based mathematics instruction, teachers on average elicited one more student contribution per hour and used one more mathematical representation per every two hours of instruction (Garet et al., 2010). Together, this body of work illustrates that teaching practice has proven surprisingly resistant and resource-intensive to change.

Our goal is to address these challenges and show that it is possible to provide consistent and effective feedback to teachers by using automated tools. Leveraging recent advances in natural language processing (NLP), we developed a tool to provide immediate feedback to teachers on their *uptake of student contributions* — namely, instances when a teacher acknowledges, revoices, and uses students’ ideas as resources in their instruction. We focus on uptake because it is a fundamental teaching skill (Collins, 1982) associated with dialogic instruction (Nystrand et al., 1997; Wells, 1999), whose positive association with student learning and achievement has been widely documented across learning contexts (Brophy, 1984; O’Connor & Michaels, 1993; Nystrand et al., 2000; Wells & Arauz, 2006; Herbel-Eisenmann et al., 2009; Demszky et al., 2021). Improving uptake has proven to be among the most difficult teaching practices to change (Cohen, 2011; Kraft & Hill, 2020) perhaps due to its cognitive complexity (Lampert, 2001). Applying our tool to a practice that has been shown difficult to alter can help demonstrate its potential to improve instruction through providing feedback to teachers.

We employed this automated tool to provide feedback to 1,136 instructors as part of Code in Place, a five-week free online computer science course organized by Stanford University. Code in Place teaches introduction to programming to ~12k students worldwide, in small sections with a 1:10 teacher-student ratio (Piech et al., 2021). This course involves a large and diverse sample of instructors and students in terms of gender, nationality and experience; they focus on the same topic and use the same language of instruction, English.

Through a randomized experiment, we demonstrate the effectiveness of our automated feedback treatment, which resulted in a 24% average increase in instructors’ uptake of student contributions. Suggestive evidence shows that this improvement in uptake is explained not by instructors’ simple repetition of student contributions but instead by more sophisticated instructional strategies such as follow-up questioning. We also find that instructors’ exposure to

our tool improves students' satisfaction with and engagement in the course. Our study creates multiple avenues for future research, including applying and extending our tool to more contexts and teaching strategies and combining automated and manual feedback in a scalable PD framework for teachers.

## **Measuring Teachers' Uptake of Student Contributions**

When teachers take up student contributions by, for example, revoicing them, elaborating on them, or asking a follow-up question, they amplify student voices and give students agency in the learning process. Given its documented positive association with student learning and achievement (Brophy, 1984; O'Connor & Michaels, 1993; Nystrand et al., 2000; Wells & Arauz, 2006; Herbel-Eisenmann et al., 2009; Demszky et al., 2021), many scholars consider uptake a core teaching strategy and an important part of classroom observation instruments. Uptake is associated with various discourse strategies (Clark & Schaefer, 1989). In education, especially effective uptake strategies include cases when a teacher follows up on a students' contribution via a question or elaboration (Collins, 1982; Nystrand et al., 1997). Repetition, for example, is considered to be a less sophisticated uptake strategy in education, but it can still serve as a way for teachers to demonstrate that they are listening to students (Tannen, 1987).

The most widely used classroom observation instruments in the U.S. such as Framework for Teaching (Danielson, 2007) and CLASS (Pianta et al., 2008) include items that measure uptake. These items, along with many others that capture similarly complex teaching strategies, are coded manually by experts through a cognitively demanding and labor-intensive process. Wells & Arauz (2006) developed an even more fine-grained hierarchical coding scheme for manually evaluating uptake. Although their scheme allows for the measurement of sophisticated uptake patterns, including various sub-categories such as follow-up questions and rejection/acceptance of student contributions, it has as many as ~230 code combinations, which makes its use too resource-intensive to scale.

Recent efforts to measure uptake at scale have sought to generate scores for this construct automatically using NLP methods. Samei et al. (2014) and Jensen et al. (2020) use automated classification to detect uptake in elementary English language arts (ELA) and math classrooms.

Their approach involves hiring experts to manually code several thousand teacher utterances for uptake (treating it as a binary variable); then, they train a machine learning classifier on the annotated utterances and apply this classifier to detect uptake in new teacher utterances. Although this approach shows promise, it requires a large number of high-quality annotations in order to train the classifier, and thus does not work well when it is not possible to obtain such annotations.<sup>1</sup> Moreover, the relationship of their measure to outcomes is yet to be explored.

In this work, we use a fully automated measure to identify uptake, which has been validated using educational outcomes across domains (Demszky et al., 2021). This measure also uses machine learning but it does not require manual annotation because it learns to identify uptake based on turn-taking patterns in classroom interaction. Specifically, the measure captures the extent to which a teacher's response is specific to the student's contribution; that connection serves as evidence that the teacher understood and is building on the student's idea (Clark & Schaefer, 1989). Demszky et al. (2021) find that this measure captures a wide range of uptake strategies, including revoicing, question answering, and elaboration, and that it correlates strongly with expert annotations for uptake. The authors also conducted a cross-domain validation and found that their measure correlates positively with instruction quality and student satisfaction across three different datasets of student-teacher interaction, including elementary math classroom transcripts, small group ELA virtual classroom transcripts and text-based math and science tutoring transcripts.

## **Providing Automated Feedback to Teachers**

Efforts to build automated feedback tools for educators are underway. Automated tools can provide teachers with objective insights on teacher practice in a scalable and consistent way and thereby offer complementary advantages to expert feedback, which is challenging to scale due to resource constraints and teachers' buy-in of inherently subjective information on their teaching (Kraft, Blazar & Hogan, 2018).

The majority of automated tools provide teachers with analytics on student engagement and progress and allow teachers to monitor student learning and intervene when needed (Alrajhi et al., 2021; Aslan et al., 2019; among others). Few tools provide teachers with feedback that can

serve as a vehicle for self-reflection and instructional improvement. To help address this gap, researchers have developed measures to detect teacher talk moves linked to dialogic instruction (Samei et al., 2014; Donnelly et al., 2017; Kelly et al., 2018; Jensen et al., 2020). For example, Kelly et al. (2018) propose an NLP measure trained on human-coded transcripts of live classroom audio to identify the number of authentic questions a teacher asks in her classroom. Moving beyond measurement to teacher feedback, Suresh et al. (2021) introduce the TalkMoves application that provides teachers with information on the extent to which they dialogic talk moves, including pressing for accuracy and revoicing student ideas. However, the impact of these tools on teacher practice is yet to be determined.

## **Our Contributions**

Our work makes two key contributions. First, we built and deployed *an automated tool that provides teachers feedback* on the extent to which they take up student contributions. This tool is reproducible and scalable because it primarily uses open-source software. In an online setting, our tool requires minimal resources because it uses a relatively low-cost automated speech recognition service and a fully automated measure for uptake that does not require annotated data. Our user interface, developed in consultation with experts in human-computer interaction and educational interventions as well as teachers themselves, is intuitive to use and non-evaluative. We share the details on the tool and the decisions we made so that researchers and practitioners can readily reproduce, build on and integrate it into their own educational platforms.

Second, to our knowledge, we are among the first to *evaluate the impact of automated feedback on teacher instruction* through a large-scale randomized study. Our study contributes to research and practice related to teachers' PD because we provide an experimental framework and first-line results in evaluating automated feedback tools for teachers. More specifically, we also contribute to the understanding of how to improve teachers' uptake, a core teaching strategy that thus far has proven difficult to change.

## Study Background

We ran the study as part of Code in Place, a 5-week-long, large-scale, free online introductory programming course organized by Stanford University (Piech et al., 2021). The mission of the course is to democratize access to teaching and learning how to code. The course was taught for the first time in Spring 2020 as a response to the COVID pandemic; due to its popularity, it was offered again in Spring 2021— which is when we conducted the experiment.

**Participants.** In Spring 2021, Code in Place enrolled ~12k students and recruited 1,136 volunteer section leaders worldwide (henceforth referred to as instructors). Instructors applied for the position by submitting both a programming exercise and a 5-minute video of themselves teaching. Each accepted instructor was assigned to teach a section with 10 students. The sections met weekly for an hour to discuss key topics in the course. The course materials were prepared in advance by the course organizers. All but nine instructors taught in English; we removed sections taught in other languages from our sample.

Sixty-five percent of our instructor sample described themselves as male, 34% as female and 1% as non-binary. Instructors ranged in age from 18-81 ( $M=29$ ,  $SD=11$ ). Instructors were located in 82 unique countries (64% in the USA, 8% in India, 3% in Canada, 2% each in Germany, Turkey and the UK, and 1% each in other countries); 21% were returning instructors from Code in Place 2020. Since the course only collected gender, age, and location information from instructors, we cannot report other demographic information (race/ethnicity, occupation, etc.). See Piech et al. (2021) for details on student demographics.

**Online setup.** The sections were taught using OhYay, an online video calling platform. Each week instructors were provided with a link for their own virtual OhYay room for meetings with their section. Instructors also had the option to use a different platform (e.g. Zoom), but in practice, 80% of the instructors used OhYay at least twice during the course. Code in Place automatically recorded each section in OhYay. All instructors consented to being recorded when choosing to use OhYay at the time they signed up for the course. We provided feedback only on sections recorded via OhYay.



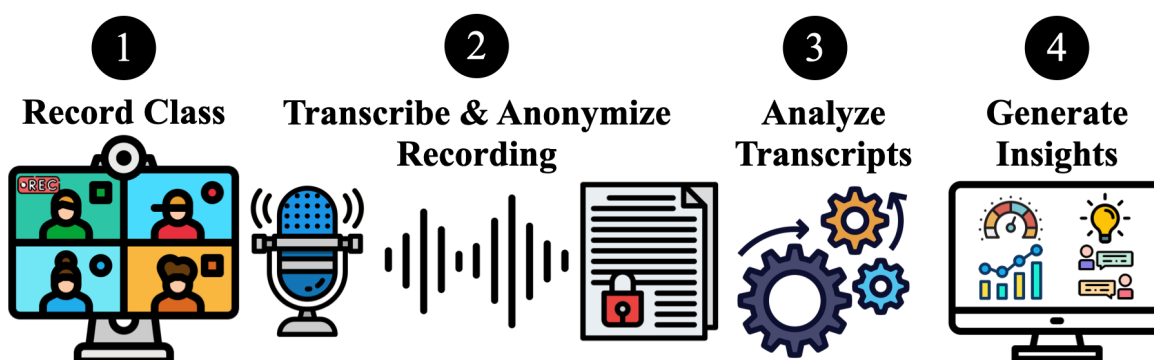
# Automated Feedback on Uptake

## Workflow for Generating Feedback

Our workflow for generating feedback is fully automated; it does not require human intervention at any step (Figure 1).

**Figure 1**

*Workflow for Generating Automated Teacher Feedback*



The workflow consists of the following steps (see Appendix A for more details):

1. **Recording.** OhYay, the video calling platform used by Code in Place, automatically records classroom verbal interactions in real time.<sup>2</sup>
2. **Transcription and anonymization.** We transcribe and algorithmically anonymize recordings using Assembly.ai, a service we chose because of its accuracy, cost-effectiveness (\$1 per 1 hr of audio) and ease of use.
3. **Transcript analysis.** We algorithmically analyze the transcripts to identify instances when a teacher takes up a student's contribution, using the measure described in Demszky et al. (2021). We also identify additional discourse features that we do not present as feedback but rather to help us break down different uptake strategies. These features include whether a teacher turn includes a question and whether a teacher repeats parts of the student utterance.
4. **Interface for teachers.** We display feedback to teachers on a web application, showing them statistics on their uptake, examples of high uptake from their transcript, and tips for improvement. Since uptake hinges on students contributing to the classroom discussion, we further facilitate teachers' interpretation of the feedback on uptake by providing

teachers with information on student engagement, including information on student talk time and examples from the transcript where the teacher's question elicited a long student response. Finally, we invite teachers to reflect on their instruction and plan for the next lesson (Appendix B). We introduce the specific design features of the feedback below.

## **Design Principles for the Automated Feedback**

Our primary objective is to encourage teachers to reflect on their practice, and thereby improve their uptake of student contributions during class sessions. To this end, we designed the automated teacher feedback tool with several principles in mind and drew on insights from experts and relevant literature in education, social psychology and human computer interaction.

We provided non-judgmental information about teachers' instruction in a way that respects their *agency* and *authority over their practice* (Wills & Haymore Sandholtz, 2009; Priestley et al., 2015; Oolbekkink-Marchand et al., 2017). Specifically, we conveyed the feedback privately to each teacher, and explicitly stated that the feedback is not used to evaluate them, but rather it is meant to support their professional development. We also included open-ended reflection questions for the teacher to elicit their own interpretation of the statistics and examples and to encourage them to give advice to themselves, following the “saying is believing” principle (Higgins & Rholes, 1978) widely recognized in social psychology.

Second, we took several steps to make the feedback *concise*, *specific* and *actionable*. With only one page of information, we used figures to visualize high-level statistics on their frequency of taking up student ideas and on student talk time. To substantiate these statistics and encourage teachers to reflect on their instruction, we highlighted examples of uptake from their transcript and asked teachers to reflect on the strategies they used in these examples. To help teachers see how their practice evolves over time and set goals for themselves, we included tabs that allowed them to revisit their feedback from earlier class sessions. We also provided advice on and examples of uptake as well as links to further resources including papers and blog posts on uptake and dialogic instruction.

Finally, we delivered the feedback in a *timely* and *regular* manner. To ensure that teachers still have a fresh memory of what they did and to make the feedback more relevant and exciting

(Shute, 2008), we shared feedback with teachers within 2-4 days after their class sessions, and always before their next class. We delivered feedback to teachers after each recorded class, with hopes that sustained work in this area would lead to improved practice over time.

## **Randomized Controlled Trial**

We conducted a randomized control trial to evaluate the effectiveness of our automated feedback tool. For ethical reasons, we offered every instructor access to the automated feedback through a link on the course website, listed as part of teaching-related resources. The key idea of our study design is to generate an exogenous variation of interacting with the feedback through encouraging a random group of teachers to read the feedback more frequently through email reminders.

Before the start of the course, we randomly assigned half of the instructors to treatment ( $n=568$ ) and the other half to control ( $n=568$ ) groups. After each of their sections, we sent instructors in the treatment group an email encouraging them to check the feedback.<sup>3</sup> In order to ensure that the intervention effect is mediated by the content of the automated feedback rather than the content of the email, we made the email short and generic, with only a link to the feedback and two non-personalized sentences encouraging instructors to take a look (Appendix C).

### **Data Collection**

**Transcripts.** The transcripts were generated automatically based on section recordings from OhYay. The course collected a total of 4,056 section recordings longer than 30 minutes<sup>4</sup>; the average duration was 64 minutes ( $SD=19$ ).

**Administrative data.** In addition to age, gender and location information for instructors introduced above, we also observed each time an instructor opened the feedback webpage and the number of students who attended each section.

**Endline survey to instructors.** We administered a survey to a randomly selected group of 200 instructors. The survey asked instructors to report their perception of the tool, the

effects this tool had on their teaching and suggestions for improving the tool (Appendix D). Instructors were sampled irrespective of treatment status, received up to three reminders and were incentivized with a chance to win one of ten \$40 Amazon gift cards. The survey achieved a 71% response rate (n=142).

**Endline survey to students.** We also administered a survey to all students without a reminder or incentive (16% response rate, n=1958). The survey asked students to report their satisfaction with the course and the helpfulness of sections (Appendix F).

All data were de-identified before analysis and linked through anonymous research IDs. Because instructors in Code in Place did not assign grades, we do not have student academic outcomes.

### Validating Randomization

To verify whether our randomization was successful, we compared treatment and control group instructor demographics. We also compared instructors' discourse features measured in their first class session, prior to receiving feedback. As Table 1 shows, we do not find statistically significant differences between conditions in any instructor demographics and discourse features in the first section. This analysis validates our randomization and suggests that any differences we observe later in the course are likely due to the effects of the intervention.

[Insert Table 1]

### Statistical models

We use the following two-stage least squares estimator (2SLS) to estimate the effects of the feedback in the context of our randomized design. We expect our randomized email intervention to increase the likelihood of the treated instructors checking the weekly feedback, which provides us the exogenous variation to estimate the causal effect of interacting with the feedback.

$$Feedback_i = \pi_0 + \pi_1 T_i + \pi_2 X_i + \varepsilon_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 \hat{Feedback}_i + \beta_2 X_i + \mu_i \quad (2)$$

In Equation (1), we model whether instructor  $i$  interacted with the feedback, — measured as a binary variable indicating whether the instructor opened the feedback page — as a function of the treatment status ( $T_i$ ) and a series of covariates ( $X_i$ ) related to instructors' demographics and pre-intervention discourse features (variables from Table 1). We then use the predicted value for feedback interaction as the predictor in the second stage (Equation 2).  $\beta_i$  is our parameter of interest that captures the local average treatment effects of our intervention. We consider several outcomes ( $Y_i$ ) to capture various aspects of instructor behavioral changes: the number of uptakes is our primary outcome as it is what the intervention is designed for, but we also consider the number of questions asked, the number of repetitions, and their talk time to further examine the mechanisms of change.

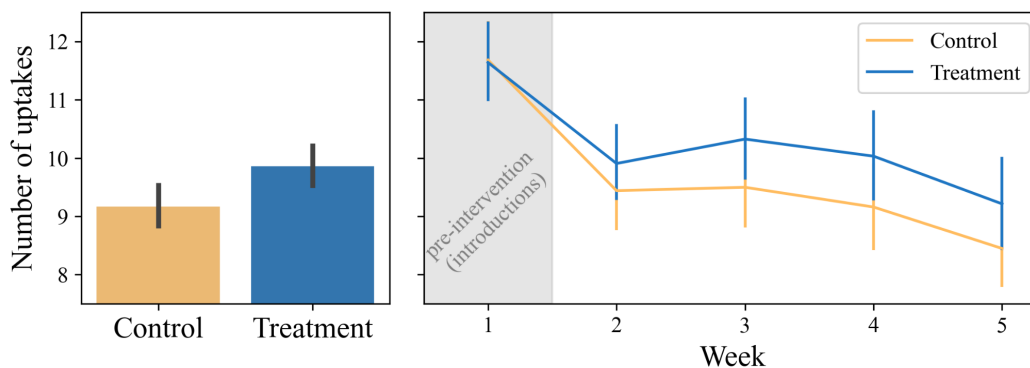
## Results

### The Impact of Our Tool on Instructors' Uptake of Student Contributions

As a first step, we conducted an intent to treat analysis (ITT) to see if there was a significant difference in the number of uptakes by condition irrespective of which instructors checked the feedback.

**Figure 2**

*Number of Times Instructors Took Up Student Contributions By Condition*



**Notes.** Each observation is a transcript, representing a unique instructor and week combination.

Figure 2 (left) shows that, compared to instructors in the control group, instructors in the treatment group took up student contributions significantly more — about one additional time on average per section, indicating a 10% increase in uptake. Figure 2 (right) shows that the

difference in uptake by condition persisted across all intervention weeks (weeks 2-5); it also shows no difference between conditions pre-intervention (week 1). The average number of uptakes irrespective of condition varied considerably across weeks (with week 1 being the highest and week 5 being the lowest) due to differences in the section focus (introductions vs reviewing material).

[Insert Table 2]

The results from our preferred 2SLS model are shown in Table 2 (Model 2). The first column presents the results for the first stage: instructors who received the email reminder were twice as likely to look at the feedback than instructors who did not receive the email ( $p < 0.001$ ); this pattern suggests that the email was effective in motivating instructors to check the feedback. Unsurprisingly, the  $F$  statistics are well above 10, suggesting our instrument (i.e., the randomization) is strong. As for the second stage, or the TOT effects, instructors who checked the feedback took up student contributions ~2.2 additional times (24%) per week ( $p < 0.05$ ), roughly 2.5 times as large as the estimate from the ITT analysis.

To help explain the increase in uptake for the treatment group, we also used discourse correlates of uptake as alternative outcomes in the second stage of the 2SLS. The correlates of uptake were the number of questions and the number of repetitions and teacher talk time, calculated based on instructors' pre-intervention transcripts.<sup>6</sup> Interestingly, we found that instructors who looked at the feedback asked roughly six (22%) more questions per class ( $p < 0.05$ ), but did not repeat student contributions more frequently nor did they talk less. These results suggest that the improvement in uptake is driven primarily by strategies other than repetition or talk time such as increased questioning.

### **Instructors' Satisfaction with the Feedback Tool**

We analyze instructors' responses to the confidential endline survey (Appendix D) to understand if they found the feedback helpful ( $n=142$ ). Instructors were strongly encouraged to report their honest opinion as a way to help improve the tool. We found that overall, instructors reported that the feedback was helpful: the majority of instructors reported that the tool 1) helped them become a better teacher (57%), 2) made them realize things about their teaching that they

otherwise would not have (76%), 3) made them pay more attention to who was getting voice in their class (57%), 4) tried new things in their teaching as a result of the feedback (53%) and that 5) the feedback wasn't difficult to understand (64%). Instructors gave an average score of 7 out of 10 for how likely they are to recommend the tool to other teachers. In the open-ended questions, the most frequently reported suggestions for improvement ( $n=62$ ) relate to improving the transcription ( $n=20$ ) and incorporating the chat into the analysis ( $n=8$ ). See Appendix E for a detailed report.

### **The Impact of Our Tool on Student Experience**

To understand if the automated feedback had an impact on students' satisfaction with the course and attendance, we fit 2SLS models using students' endline survey responses ( $n=1958$ ) and attendance data. As Table 3 (Model 2) shows, students who were assigned to instructors who checked the feedback were significantly more likely to respond to the survey ( $p<0.05$ ), recommend the class ( $p<0.05$ ) and find the sections to be helpful ( $p<0.05$ ).<sup>7</sup> We did not observe a significant difference in student attendance based on whether instructors checked the feedback.

[Insert Table 3]

## **Discussion**

Our study investigated whether it is possible to deliver feedback to teachers at scale effectively using automated tools. We developed a fully automated tool to provide feedback to teachers on their uptake of student contributions, one of the most important discourse phenomena associated with dialogic instruction, and to test the effectiveness of this tool in a large-scale online programming course. In doing so, we demonstrated that feedback on instruction, typically a labor-intensive process that often meets significant resistance from teachers, can be delivered widely and can stimulate improvements in instructional practice.

We found that the automated teaching insights in our tool increased instructors' uptake of student contributions by 24%, a result likely driven by instructors' increased use of more

sophisticated strategies beyond repetition such as questioning. Our analyses of survey responses provided evidence that the majority of instructors found the feedback helpful. There is also suggestive evidence that students whose teachers looked at the feedback more frequently were more satisfied with and engaged in the course. These results together suggest that our tool has a positive impact on instruction. Furthermore, the fact that we were able to improve a phenomenon as complex as teacher uptake using automated feedback indicates the potential for improving other teaching strategies. However, there are also limitations to the current study. Addressing these limitations can serve as an important step towards exploring the full potential of automated tools for teachers.

Our study took place in an online programming course where many instructors are novices. We focused on only one fundamental teaching practice: teachers' uptake of student ideas. Thus, our automated feedback approach requires a series of follow-up studies to test whether the results can hold for other teaching practices and in educational settings with different parameters regarding course subjects, teachers' experience level and composition of students. Applying our approach to a setting where student learning outcomes are available would also help determine whether the improvement in teaching practice induced by the automated feedback translates into improvements in students' academic achievement.

Our study has technological limitations that need to be addressed in future research as well. For example, our tool relies on an automated speech recognition service, which is less accurate for speakers whose native language is not Standard American English. Differences in speech recognition accuracy based on teacher and student demographics are problematic because they may continue to propagate inequities in teachers' PD. We sought to be conscious about this issue by emphasizing to participants that we were conducting a pilot study and we were at a nascent stage of testing this tool. We plan to address speech recognition issues by leveraging technological improvements in this area that mitigate biases and by using custom models trained and evaluated on audio data representative of teachers and students.

Additionally, as of now the tool can only analyze spoken English conversations between the teacher and students. Since the NLP-based measure for uptake does not require manual annotation, it is possible to extend the tool to other languages where an automated transcription



service and a dataset of classroom interactions are available. Including other communication pathways such as chat messages and video would allow the tool to capture important aspects of online instruction beyond speech.

Despite its limitations, this study constitutes an important step towards our ultimate goal of developing an effective, scalable feedback tool for all teachers. With the development of new NLP-based measures of instruction, we can extend our tool to generate insights on multiple aspects of teaching (Liu & Cohen, 2021). While building the technological setup to record in-person classrooms requires substantial initial investment (e.g., Kelly et al., 2018; Jensen et al., 2020), applying our tool in K-12 settings offers particular promise as K-12 teachers have been proven to be the most influential within-school factor for student learning and life outcomes (Chetty, Friedman, & Rockoff, 2014). Besides providing information to teachers directly, our automated tool might also complement existing PD efforts by assisting coaches in observing and evaluating instruction and letting coaches spend more time having individualized, evidence-based, improvement-focused conversations with teachers. Future efforts should continue to improve, validate and apply the automated feedback tool studied here to explore its full potential to support teaching and improve student learning outcomes across educational contexts.

## Notes

<sup>1</sup> Jensen al. (2020) actually removed uptake from their analysis because it occurred too infrequently in their annotated data.

<sup>2</sup> Since our focus is whole class interaction, we record the main class exclusively, and do not record breakout rooms. This decision does not affect our data significantly, as in our case teachers spend only 1% of class time in breakout rooms, likely due to the small class size.

<sup>3</sup> The feedback (and hence, the email) was available if they taught their section on OhYay themselves (did not have substitutes).

<sup>4</sup> We removed recordings shorter than 30 minutes, which indicated that the instructor did not use OhYay for their entire lesson.

<sup>5</sup> This lack of difference holds even if we include all categories for gender and for country rather than using binary categories.

<sup>6</sup> We calculated correlations based on transcripts from the first week (pre-intervention), by regressing the number of instructor uptakes on each discourse variable while controlling for session duration. The standardized coefficients for instructors' discourse features are: number of questions ( $\beta=0.878, p < 0.001$ ), number of repetitions ( $\beta=0.824, p < 0.001$ ) and talk time in minutes ( $\beta=-0.716, p < 0.001$ ).

<sup>7</sup> We do not have a reason to believe that these differences are due to teachers in the treatment group directly telling students to respond to the survey, since teachers were not aware of the intervention and most of them were also not aware of student endline surveys. Thus, we can reasonably assume that these differences are due to an indirect effect of teacher practice on student satisfaction.

## References

- Alrajhi, L., Alamri, A., Pereira, F. D., & Cristea, A. I. (2021, June). Urgency Analysis of Learners' Comments: An Automated Intervention Priority Model for MOOC. In *International Conference on Intelligent Tutoring Systems* (pp. 148-160). Springer, Cham.
- Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Arslan Esme, A. (2019, May). Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-12).
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. *Teaching as the learning profession: Handbook of policy and practice, 1*, 3-22.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher, 33*(8), 3-15.
- Brophy, J. E. (1984). Teacher behavior and student achievement (No. 73). Institute for Research on Teaching, Michigan State University.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American economic review, 104*(9), 2633-79.

- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, 13(2), 259-294.
- Cohen, D. K. (2011). *Teaching and its predicaments*. Harvard University Press.
- Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378-387
- Collins, J. (1982). Discourse style, classroom interaction and differential treatment. *Journal of Reading Behavior*, 14, 429-437.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. ASCD.
- Demszky, D., Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., & Hashimoto, T. (2021). Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP)*.
- Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017). Words matter: Automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 218–227.  
<https://doi.org/10.1145/3027385.3027417>
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., ... & Szejnberg, L. (2008). The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement. NCEE 2008-4030. *National Center for Education Evaluation and Regional Assistance*.
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., ... & Doolittle, F. (2010). Middle School Mathematics Professional Development Impact Study: Findings After the First Year of Implementation. NCEE 2010-4009. *National Center for Education Evaluation and Regional Assistance*.
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., ... & Doolittle, F. (2011). Middle School Mathematics Professional Development Impact Study: Findings after the Second Year of Implementation. NCEE 2011-4024. *National Center for Education Evaluation and Regional Assistance*.

- Gregory, A., Ruzek, E., Hafen, C.A., Mikami, A.Y., Allen, J.P. and Pianta, R.C., 2017. My Teaching Partner-Secondary: A video-based coaching model. *Theory into practice*, 56(1), pp.38-45.
- Herbel-Eisenmann, B., Drake, C., & Cirillo, M. (2009). "Muddying the clear waters": Teachers' take-up of the linguistic idea of revoicing. *Teaching and Teacher Education*, 25(2), 268-277.
- Hill, H. C., & Erickson, A. (2019). Using Implementation Fidelity to Aid in Interpreting Program Impacts: A Brief Review. *Educational Researcher*, 48(9), 590–598.  
<https://doi.org/10.3102/0013189X19891436>
- Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, 14(4), 363-378.
- Jacob, A., & McGovern, K. (2015). The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development. *TNTP*.
- Jensen, E., Dale, M., Donnelly, P. J., Stone, C., Kelly, S., Godley, A., & D'Mello, S. K. (2020). Toward Automated Feedback on Teacher Discourse to Enhance Teacher Learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3313831.3376418>
- Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically Measuring Question Authenticity in Real-World Classrooms. *Educational Researcher*, 47(7), 451–464. <https://doi.org/10.3102/0013189X18785613>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Kraft, M. A., & Hill, H. C. (2020). Developing ambitious mathematics instruction through web-based coaching: A randomized field trial. *American Educational Research Journal*, 57(6), 2378-2414.
- Lampert, M. (2001). Teaching problems and the problems of teaching. Yale University Press.

- Liu, J., & Cohen, J. (2021). Measuring Teaching Practices at Scale: A Novel Application of Text-as-Data Methods. *Educational Evaluation and Policy Analysis*, 01623737211009267.
- Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260-293.
- Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue* (pp. 30-61). New York: Teachers College Press.
- Nystrand, M., Wu, L.L., Gamoran, A., Zeiser, S. and Long, D. (2000) Questions In time: Investigating the structure and dynamics of unfolding classroom discourse. National Research Center on English Learning and Achievement (CELA).The University of Wisconsin-Madison.
- O'Connor, M. C., & Michaels, S. (1993). Aligning Academic Task and Participation Status through Revoicing: Analysis of a Classroom Discourse Strategy. *Anthropology & Education Quarterly*, 318-335.
- Oolbekkink-Marchand, H. W., Hadar, L. L., Smith, K., Helleve, I., & Ulvik, M. (2017). Teachers' perceived professional space and their agency. *Teaching and teacher education*, 62, 37-46.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H Brookes Publishing.
- Piech, C., Malik, A., Jue, K., & Sahami, M. (2021, March). Code in Place: Online Section Leading for Scalable Human-Centered Learning. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 973-979).
- Priestley, M., Biesta, G. J. J., Philippou, S., & Robinson, S. (2015). The teacher and the curriculum: Exploring teacher agency. *The SAGE handbook of curriculum, pedagogy and assessment*, 187-201.
- Russell, J. L., Correnti, R., Stein, M. K., Bill, V., Hannan, M., Schwartz, N., ... & Matthis, C. (2020). Learning from adaptation to support instructional improvement at scale:

- Understanding coach adaptation in the TN Mathematics Coaching Project. *American Educational Research Journal*, 57(1), 148-187.
- Samei, B., Olney, A. M., Kelly, S., Nystrand, M., D'Mello, S., Blanchard, N., Sun, X., Glaus, M., & Graesser, A. (2014). *Domain Independent Assessment of Dialogic Properties of Classroom Discourse*. <https://eric.ed.gov/?id=ED566380>
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, 10(4), 535-572.
- Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J. H., & Sumner, T. (2021). Using Transformers to Provide Teachers with Personalized Feedback on their Classroom Discourse: The TalkMoves Application. *arXiv preprint arXiv:2105.07949*.
- Tannen, D. (1987). Repetition in conversation: Toward a poetics of talk. *Language*, 574-605.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628-51.
- Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge University Press.
- Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *The journal of the learning sciences*, 15(3), 379-428.
- Wills, J. S., & Haymore Sandholtz, J. (2009). Constrained professionalism: Dilemmas of teaching in the face of test-based accountability. *Teachers college record*, 111(4), 1065-1114.

**Table 1*****Descriptive Statistics of Teacher-Level Variables, Verifying Randomization***

<b>Variable</b>	<b>Control, n=568 n (%) / M (SD)</b>	<b>Treatment, n=568 n (%) / M (SD)</b>	<b>t-value</b>	<b>p-value</b>
<b><i>Demographics</i></b>				
Female	186 (32%)	192 (33%)	0.38	0.71
Age	28.5 (10.5)	29.0 (11.4)	1.64	0.10
In USA	361 (63%)	363 (63%)	0.12	0.90
Returning teacher	113 (19%)	124 (21%)	0.80	0.42
<b><i>Pre-intervention statistics (week 1)</i></b>				
Transcripts available	465 (81%)	480 (84%)	1.19	0.23
Section duration (min)	63.9 (16.2)	65 (19.5)	1.23	0.22
Student attendance (n)	5.7 (2.1)	5.0 (2.1)	0.50	0.62
Students speaking (n)	4.8 (2.2)	4.0 (2.2)	0.70	0.70
<b><i>Teacher practices</i></b>				
Uptakes (n)	11.7 (7.0)	11 (7.2)	-0.09	0.93
Questions (n)	33.8 (16.7)	34.0 (18.0)	0.49	0.63
Repetitions (n)	36.3 (18.8)	36.0 (19.2)	0.21	0.83
Talk time (min)	48.6 (15.4)	50.0 (18.2)	1.48	0.14

**Notes.** The randomization was performed prior to the start of the course (before week 1).

**Table 2*****Effects of Automated Feedback on Teacher Practices***

	1st Stage	2nd Stage (Teacher Practices)			
	(1)	(2)	(3)	(4)	(5)
Independent variable	Checked feedback	Uptakes (n)	Questions (n)	Repetitions (n)	Talktime (min)
Control mean	0.219	9.056	29.148	33.569	50.971
<i>Model 1 (n=880)</i>					
Checked feedback	N/A	2.306 <sup>†</sup> (1.291)	6.771 <sup>†</sup> (3.593)	4.782 (4.120)	-2.599 (2.034)
Email reminder	0.277*** (0.024)	N/A	N/A	N/A	N/A
F stat. / Adj. R <sup>2</sup>	32	0.132	0.159	0.212	0.701
<i>Model 2 (+teacher-level covariates, n=879)</i>					
Checked feedback	N/A	2.209* (1.070)	6.210* (2.882)	4.355 (3.478)	-2.512 (1.835)
Email reminder	0.278*** (0.023)	N/A	N/A	N/A	N/A
F stat. / Adj. R <sup>2</sup>	16	0.408	0.462	0.442	0.758

**Notes.** Each column comes from a separate regression. The sample includes all teachers who showed up in the first week and taught at least another session in week 2 to 5. The number of weeks a teacher taught post week 1 is not affected by treatment status ( $\beta = -0.6$ ,  $p = 0.284$ ). For columns (2) to (5), all the outcome measures are averages of a teacher's practice from week 2 to 5. In both models, we control for class duration (min) and binary variables for each week indicating whether the teacher had a transcript that week. For Model 2, we include teacher-level covariates, including gender, whether a teacher is from the USA, age, whether a teacher is a returning teacher) and teacher practices in week-1 session (number of uptakes, number of repetitions, number of questions, talktime in minutes). \* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .



**Table 3*****Effects of Automated Feedback on Student Evaluation and Engagement***

	<b>2SLS Estimates (2nd Stage)</b>			
	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
<b>Independent variable</b>	<b>Percentage of students responding to survey</b>	<b>Percentage of students recommending the course</b>	<b>Percentage of students rating the section as helpful</b>	<b>Average student attendance</b>
Control mean	0.152	0.142	0.080	3.84
<i>Model 1 (880)</i>				
Checked feedback	0.080** (0.029)	0.088** (0.029)	0.046* (0.022)	0.698 (0.395)
<i>Model 2 (+teacher-level covariates, n=879)</i>				
Checked feedback	0.069* (0.029)	0.078* (0.029)	0.046* (0.022)	0.364 (0.364)

**Notes.** Each column comes from a separate regression. The sample includes all students assigned to teachers who showed up in the first week and taught at least another session in week 2 to 5. The analysis is conducted at the teacher level. To compute the outcome variables for columns (1) to (3), we use the percent of students who responded to the survey, who recommended the course in the survey, and who rated the course as useful, respectively. The denominator is the total number of students assigned to a given teacher. We code a student to have recommended the course with a rating of 7 or above out of 10. We code a student to have found the section helpful if they selected either “somewhat helpful” or “very helpful” as a response (see survey in Appendix F). Student attendance is measured as the average number of students attending sections post week 1. \* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .



# Supplementary Material

## Appendix A

### Transcription & Anonymization

Since OhYay does not have built in automated transcription, we experimented with multiple transcription services and chose Assembly.ai, as we found it to be the best in terms of accuracy, cost-effectiveness (\$1 per 1 hr of audio) and ease of use (via a simple API). Speaker separation (also referred to as diarization) is available in Assembly.ai, but since we were unsure about its accuracy, we perform our own diarization by aligning speaker timestamps obtained from OhYay with word-level timestamps obtained from Assembly.ai. Although we do not expect our transcripts to contain any sensitive data, to be careful we anonymize transcripts automatically via Assembly.ai by redacting all words that could potentially refer to people, organizations, locations, phone numbers or credit card numbers. We also replace all speaker IDs with identifiers such as “Teacher”, “Student 1”, “Student 2”, etc..

### Transcript Analysis

We algorithmically analyze the transcripts to identify various discourse-related phenomena that the feedback is based on. We provide details on each of these below.

**Student and teacher talk time.** We quantify teacher and student talk time using timestamps from the transcripts. Specifically, we sum up the duration of each teacher utterance and compute talk time in minutes for our analyses.

**Number of unique students speaking in class.** Since we are able to separate speakers, we can readily obtain the number of unique students that spoke in each class.

**Teacher questions.** We build a question detector to identify teacher questions. The question detector flags an utterance as containing a question either if 1) it contains a question mark, or 2) if our NLP model identifies a question in it, since punctuation from Assembly.ai may not always be accurate. We develop this NLP model using Switchboard (Godfrey et al., 1992), a large corpus of manually transcribed phone conversations that is used often for dialog-related analyses in NLP. We strip all question marks from Switchboard and use those question marks as labels to fine-tune BERT (Devlin et al., 2019), a state-of-the-art NLP model to predict the presence of

question marks based on the utterances that are stripped of question marks. This model achieves an accuracy above 90%, and hence we rely on it to catch potential false negatives for teacher questions that we could not detect by purely checking for question marks in our transcripts.

**Teacher repetition.** We use the %-IN-T measure from Demszky et al. (2021) to detect instances where the teacher repeats parts of the student utterance. This measure computes the percentage of student words that are part of the teacher utterances, ignoring stopwords and punctuation. We identify stopwords using NLTK's list of stopwords for English (Bird, 2006).

**Teacher uptake.** We identify whether a teacher takes up a student's contribution using the automated measure described in Demszky et al. (2021), who call their measure JSD, short for Jensen Shannon Divergence. We consider any score greater than 0.8 as an example of uptake, which is a threshold we set based on the binomial distribution of scores (0.8 is the split between the two normal distributions) and based on manual inspection.

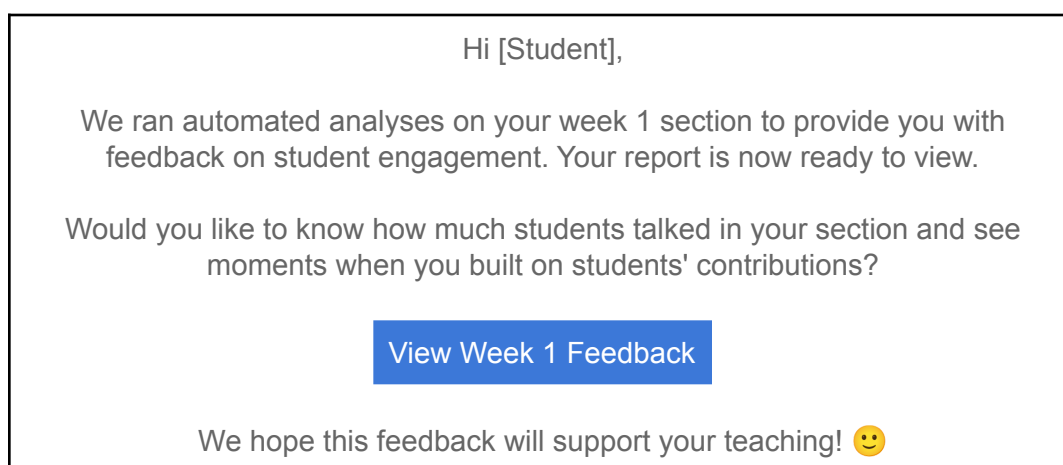
## Appendix B

Please see attached pdf of the user interface.

## Appendix C

**Figure 4**

*Email Reminder (Treatment)*



## Appendix D

### Final Survey For Instructors

We shared the following final survey about the automated feedback tool with a randomly selected sample of 200 instructors. To encourage a high response rate, these instructors received the incentive of a chance to win one of ten \$40 Amazon gift cards and we also sent 3 email reminders about the survey.

### Figure 5

*Final Survey for Instructors About the Automated Feedback*

### AI-Based Feedback on Your Week 1 Section


#### Demo

At Code in Place, we believe in the power of collaborative learning, which has also been shown to lead to student success.

Powered by state of the art AI, we provide you with feedback on two key mechanisms of student engagement: student talktime and moments when you built on student contributions.

This feedback is meant to give you an opportunity to reflect and to support your professional development. It is not meant as an evaluation.

**Notes:** 20% of your section was spent in breakout rooms, which are not analyzed here. Our language-based algorithms right now only work for sections taught in English.



### Transcript Feedback Survey

---

The Transcript Feedback component of Code in Place was part of a pilot research project. The goal of this project is to understand the usefulness of AI-powered transcript feedback to teachers like you. Thus, your feedback is essential to our project. 😊

We are looking for honest feedback, which will help us decide if we should use this tool again and how we can improve it if we do. Your responses are confidential: they will never be linked with your name (only with an anonymous research ID) and they will never be shared or used in any way to reveal your identity, not even to researchers on the Code in Place team.

**How often did you engage with the Transcript Feedback?**  
*Select one response.*

- Not at all.
- Once or twice.
- Regularly (most weeks).

---

*If they selected "Not at all":*

**Could you tell us why you didn't engage with the Transcript Feedback?**

*Select all that apply*

- I didn't know about it.
- It wasn't available to me (e.g. I didn't use Ohyay / my section wasn't in English / I had substitute section leaders).
- I didn't have the time.
- I didn't think it would be helpful.
- Other (*please explain*)

Submit

---

*If they selected "Once or twice" :*

**Could you tell us why you engaged with the Transcript Feedback only once or twice?**

*Select all that apply*

- I only learned about it later in the course.
- It wasn't available to me after each section (e.g. I didn't use Ohyay / my section wasn't in English / I had substitute section leaders).
- I didn't have the time.
- I didn't find it helpful.
- Other (*please explain*)

---

*If they selected "Once or twice" or "Regularly most weeks"):*

**To what extent do you agree with the following about the Transcript Feedback?**

*Please select one option for each: "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly agree".*

- The feedback has helped me become a better teacher.
- The feedback made me realize things about my teaching that I otherwise would not have.
- The feedback was difficult to understand.
- The feedback made me pay more attention to who was getting a voice in my class than I otherwise would have.
- I tried new things in my teaching because of this feedback

**On a scale from 0-10, how likely are you to recommend the Transcript Feedback tool to other teachers?**

*Please select between 0-10*

---

### AI-Based Feedback on Your Section

**Ability to compare to previous weeks**

**Talktime percentage**

**Number of times you built on student contributions**

**Class average**

**Examples from transcript**

**Teaching advice (with strategies and examples)**

**Reflection questions**

**Resources**

**Resources**

**Please select the MOST helpful elements of the feedback.**

*Please select between 0-3 elements*

- Ability to compare to previous weeks
- Talktime percentage
- Number of times you built on student contributions
- Class average for talktime
- Examples from your transcript for things you said that got students to talk
- Examples from your transcript for moments when you built on student contributions
- Teaching advice (with strategies and examples)
- Reflection questions
- Resources
- Other (*please explain*)

**Please select the LEAST helpful elements of the feedback.**

*Please select between 0-3 elements*

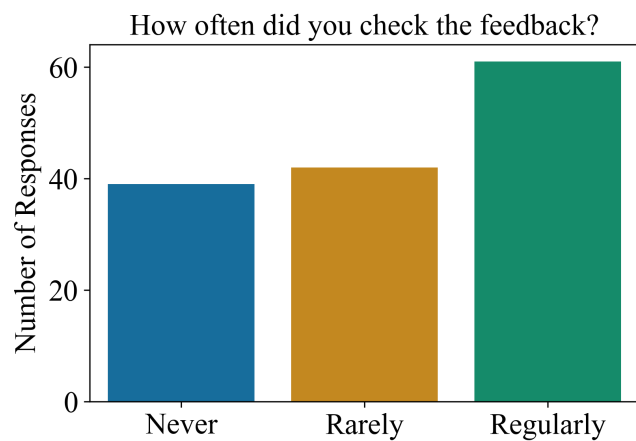
- Ability to compare to previous weeks
- Talktime percentage
- Number of times you built on student contributions
- Class average for talktime
- Examples from your transcript for things you said that got students to talk
- Examples from your transcript for moments when you built on student contributions
- Teaching advice (with strategies and examples)
- Reflection questions
- Resources
- Other (*please explain*)

**Do you have any suggestions for how we could improve this feedback tool? (open ended response)**

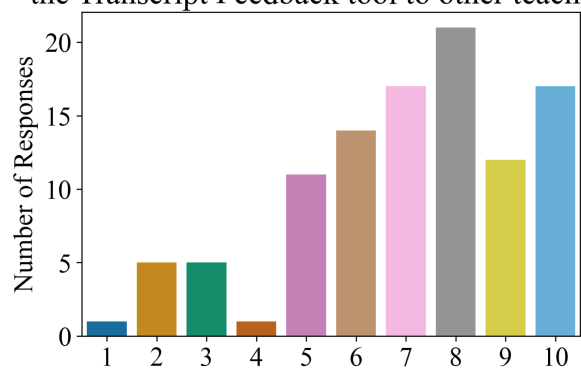
**Do you have any other thoughts / comments? :) (open ended response)**

Submit

## Appendix E

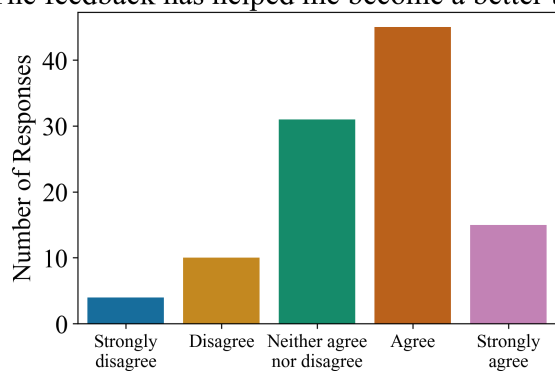


On a scale from 0-10, how likely are you to recommend the Transcript Feedback tool to other teachers?

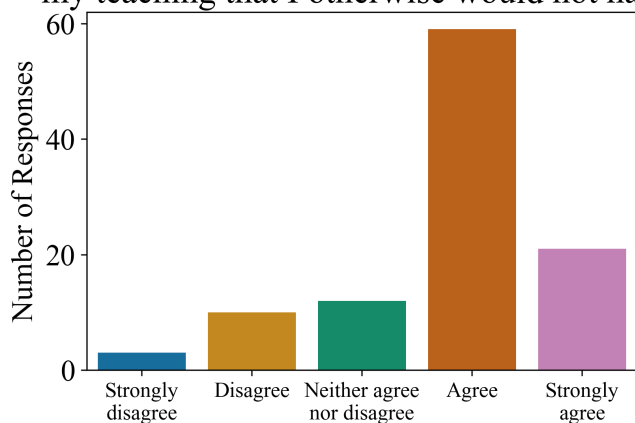




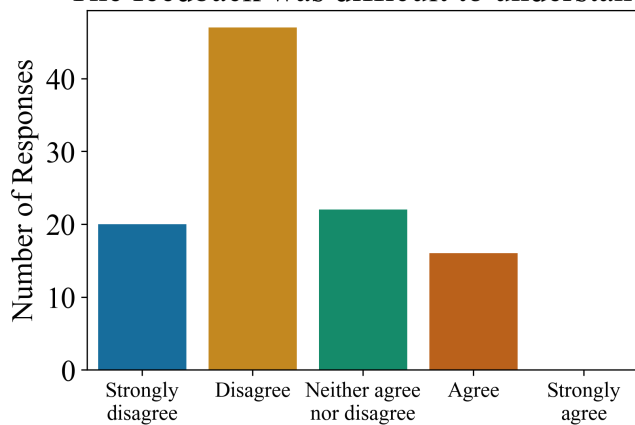
The feedback has helped me become a better teacher.



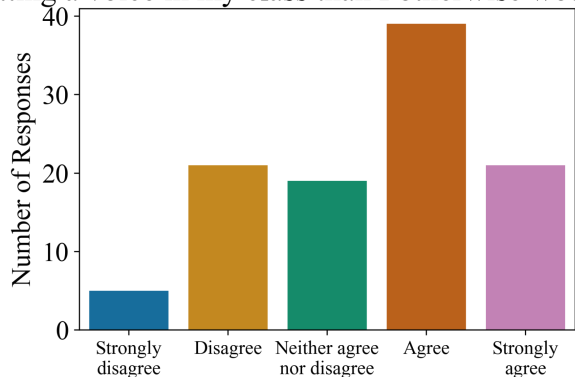
The feedback made me realize things about my teaching that I otherwise would not have.



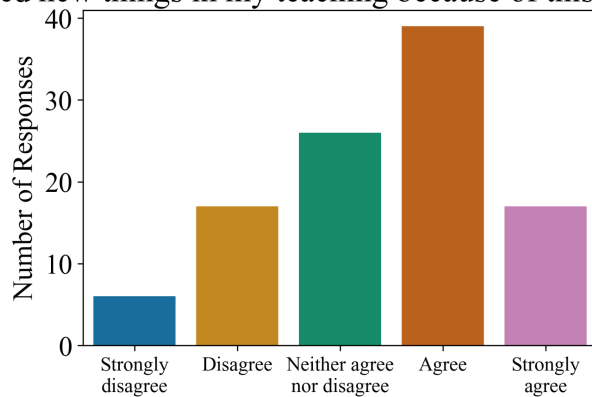
The feedback was difficult to understand.



The feedback made me pay more attention to who was getting a voice in my class than I otherwise would have.



I tried new things in my teaching because of this feedback.



## Appendix F

**Figure 6**

*Final Survey for Students About the Course*

### Code in Place Survey

We truly appreciate that you took time for Code in Place. It has been so wonderful to go on this adventure of a course with you.

Now that we're wrapping up, we'd like to ask you for a very short reflection on your time with Code in Place. We are always working on improving our own teaching, and the experience we provide students. Filling out this anonymous feedback form will help us decide if we should do this again and how we can improve it if we do.

**1. What did you like about Code in Place?**

**2. What would you improve about Code in Place?**

**3. On a scale from 0-10, how likely are you to recommend being a student in Code in Place to a friend who wants to learn to program?**

**4. Which of these course elements were helpful?**

*Please select one option for each: "Did not use", "Not very helpful", "Somewhat helpful", "Very helpful".*

- Course lectures
- Small group sections
- Ed discussion forum
- Course Assignments
- Worked Examples

**5. Leave a message for a student thinking of applying to Code in Place!**

**Have a story to tell? Email us!**

If you feel like something exceptionally positive happened to you that you would like to highlight, please do email [codeinplacestaff@gmail.com](mailto:codeinplacestaff@gmail.com)

Submit

## Appendix G

On a scale from 0-10, how likely are you to recommend being a student in Code in Place to a friend who wants to learn to program?

